

Two-sided network effects, bank interchange fees, and the allocation of fixed costs

Mats A. Bergman
Swedish Competition Authority
and
Department of Economics, Stockholm University,
SE-106 91 Stockholm, Sweden
Email: mats.bergman@ne.su.se and mats.bergman@kkv.se

January 3, 2005

Abstract

Two-sided network effects in card payment systems are analysed under different market structures, e.g., competition, one-sided monopoly, bilateral monopoly and duopoly; with and without an interchange fee; for the so-called Baxter case of non-strategic merchants. A partial ranking of market structures according to their welfare effects is provided.

Fixed central (card) system costs are introduced and analysed under free entry and duopoly. It is shown that under free entry, a per-transaction distribution of fixed costs is preferable to dividing the fixed costs in equal proportions between the participants. Under duopoly, (and no entry) a fixed division of central costs will yield lower prices.

Keywords: Two-sided markets, card payments, payment systems, acquiring, issuing, market structure

JEL code: G21, L11, L44

Acknowledgements: I am grateful for comments from Gabriela Guibourg and Björn Segendorff, as well as for financial support from the Swedish Riksbank.

1 Introduction

Payment systems, such as credit-card networks, have features that make economic analysis particularly challenging, for regulators as well as for academic economists. In order to process a transaction between customers of different banks, the banks need to cooperate extensively. Sometimes, this means accessing each others' facilities or customers. Cooperation may also be necessary in order to reap economies of scale. For example, central parts of the payment system will often provide services to many banks and may be jointly owned by its customers. In addition, in order to achieve optimal network effects, a fee structure that effectively taxes one type of final customers (e.g., merchants that accept card payments) in order to subsidize another category of customers (e.g., cardholders) may have to be implemented. The latter, in turn, may necessitate a multilateral pricing agreement between the banks.

In general, regulatory authorities have well-founded reasons to be suspicious of close cooperation between competitors, in particular concerning pricing decisions. Competition between independent firms will foster efficiency and tends to bring prices down to costs. In the provision of payment services, however, completely independent competition will not be a feasible alternative. The best available option may be competition at the retail level in combination with cooperation at the upstream (system) level. On the other hand, close cooperation at one level may give the banks the opportunity to cleverly design system fees and multilateral fees in such a way that downstream collusion is induced. This could, for example, be achieved by raising the appropriate marginal costs, so that incentives are created for the banks to raise final-customer prices and so that excess profits are generated elsewhere in the system. Considerations of this type has drawn the attention of regulators, resulting in a substantial amount of regulatory activity. The EU Commission, for example, forced Visa and Mastercard to reduce their multilateral interchange fees (see below for a discussion of interchange fees; see Bergman, 2003 and Chakravorti, 2003 for references to cases).

For much the same reasons, payment systems have spawned a considerable economics literature in recent years. In a wider context, these contributions can be seen as a part of the growing literature on two-way network effects (Rochet and Tirole, 2003; Armstrong, 2004). A key insight of this literature is that network effects between two different types of consumers have to be analysed in a system-wide context.

A specific example of a market with two-sided network effects is the payment card market (the market for debit and credit cards). Such markets have four types of participants, as shown in Figure 1: cardholders (or customers), issuers, acquirers and merchants. Cardholders use cards to pay for goods and services provided by merchants. The cards are provided by issuing banks, while acquiring banks provide services to merchants. When a card payment is registered with a merchant, the acquiring bank charges the issuing bank, which in turn charges the cardholder's account. For its services, the acquiring bank subtracts a fraction, the so-called merchant fee, from the payment to the merchant's ac-

Multilateral interchange fees

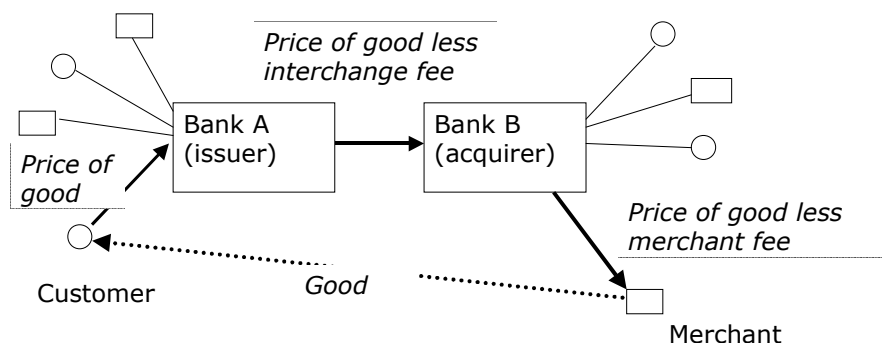


Figure 1:

count. Typically, the acquiring bank will have to pay in interchange fee to the issuing bank, the so-called interchange fee, which is therefore subtracted from the payment from the issuer to the acquirer. The merchant fee and the interchange fee can be either a percentage of the transaction value, or a fixed fee. In most cases, the cardholder pays an annual fee, while in some cases, he or she will (also) have to pay per-transaction fees. Credit cards may be free of charge for the cardholder, in particular in the USA, as long as the accumulated debt is paid monthly.

Relative to cash payments, card payments potentially generate benefits for customers, as well as for merchants. For cardholders, these benefits will increase with the number of merchants that accept cards, while merchants will be more willing to accept a card that has a larger base of cardholders. This interdependence creates two-way network effects, which will be discussed more extensively in Section 2. In the presence of network effects, the competitive equilibrium may be inefficient.

The interchange fee, which is set multilaterally by the banks, provides an instrument for improving efficiency. A higher interchange fee tends to reduce cardholders' fees and to increase the merchant fee. The first-generation litera-

ture on two-way network effects in payment markets (Baxter, 1983, Schmalensee, 2002) analysed these issues under the assumption that cardholders' and merchants' demand for card-payment services reflect the intrinsic benefits - such as convenience and safety - that they derive from card payments (relative to cash payments). A typical result is that there will be underprovision of card services, because private agents do not internalize positive network effects that accrue to other agents.

An important insight, however, was that merchants with market power (with a positive price-cost margin) may have *strategic* reasons to accept cards. By accepting cards, they will attract customers away from other merchants that do not accept cards. Conversely, if they do not accept cards, they may lose customers to other merchants that do. These strategic motives do not correspond to social gains. It follows that there may potentially be *overprovision* of card services (Katz, 2001). In the second-generation literature (e.g., Rochet and Tirole, 2002, 2003), these strategic motives are explicitly accounted for by incorporating merchant competition in the models (see further Section 10).

This paper returns to the first-generation modelling assumptions. Although possibly important, merchant market power greatly complicates the analysis: there may still be insights to be drawn from a simpler modelling approach. In addition, merchant market power may not always be significant or, perhaps more plausibly, consumers may be uninformed as to whether the merchant accepts card payments or not when they choose which store to patronize. A specific reason to go back to a simpler model is to instead introduce fixed system costs. Typically, the literature assumes that there are only marginal costs associated with payment systems, while in practice there will be substantial fixed costs. For example, there will be fixed development costs and some of the costs for setting up equipment for communication, central processing and data storage will be fixed. From a policy point-of-view, the allocation of fixed costs is an important issue, possibly with substantial implications for the barriers-to-entry into the industry.

The outline of the paper is as follows. Section 3 sets up the model. In Sections 4 through 8 the basic model is applied to different market structures: perfect competition and welfare maximum (Section 4), second-best regulation of the interchange fee (Section 5), bilateral monopoly (Section 6), a proprietary system (Section 7) and a one-sided monopoly (Section 8). Section 9 makes welfare comparisons between these market configurations. In section 10, some aspects of oligopoly interaction are introduced and further references to the literature are provided. In section 11, central system fees are introduced. After that, the effect of system fees are analysed, under the additional assumption of free entry (Section 12) and no-entry Cournot competition (Section 13). Finally, section 14 concludes.

2 Network effects

In a market without network effects, the consumer cares only about his or her own level of consumption and the price. In a market with network effects, the consumer cares - directly or indirectly - also for other consumers' levels of consumption. In the simplest setting, the number of other consumers of the same product has a direct effect on the (marginal) utility of consuming a unit of the product. For example, a given consumer's utility from having a phone or a fax increases with the number of other consumers that also have phones and faxes, respectively. This type of network effect is sometimes called a one-sided network effect.

A somewhat more complex situation arises when there are two types of agents that interact on one "platform". Either type cares for the number of agents of the *other* type that uses the platform, but not (directly) about the number of users of its own type. Some examples are buyers and sellers in advertising markets and marketplaces for trading (e.g., stock markets), as well as matchmaking markets (such as dating agencies, real estate agents and business-to-business websites). A buyer does not benefit from the presence of other buyers - and may indeed suffer from the increased competition for the sellers' product that additional buyers bring. On the other hand, the buyer derives benefit from the presence of additional sellers, while the sellers derive benefit from the presence of additional buyers. Hence, buyers may indirectly benefit from there being a large number of other buyers, as this will attract a large number of sellers - and vice versa. This phenomenon is known as two-sided network effect. Another example is the market for operative systems for personal computers: the operative system is a platform that is used by software manufacturers and by users of personal computers. An operative system such as Windows, that has a large installed base of users, is an attractive platform for software developers. Conversely, if a large number of applications have been developed for an operative system, that system will be attractive for new users. More generally, many manufacturing standards (computers and peripherals, CD players and CDs, et cetera) and communication protocols are examples of markets with two-sided network effects. Yet another example is shopping malls, which must attract customers as well as retailers.

In the financial markets, a payment-card system is an example of a two-way market: cardholders cannot interact with other cardholders and nor can merchants interact with other merchants (e.g., using their EFTPOS terminals), but cardholders can interact with merchants. Although, for some purposes, it may be appropriate to view a giro system as a two-sided market, with business and non-business customers, a giro system is perhaps better thought of as a market with one-sided network effects. Some account holders (non-business customers) may not be able to receive payments, but all account holders (business customers as well as non-business customers) are able to make payments.

Sometimes a third type of network effect is identified: indirect network effects in one-sided markets. Possible examples are public-transport networks and (single-bank) ATM networks. A higher number of passengers and a higher

number of cardholders on the ATM network, respectively, will result in more frequent departures and a denser (or wider) ATM network. This increases welfare for the average customer, even though congestion effects may imply that the *direct* effect on a given passenger's utility of another passenger may be negative, and similarly for an additional ATM cardholder. This type of network effect is very reminiscent of ordinary scale (or density) economies: as the number of customers in a retail outlet increases, the retailer can expand its range of products, it can extend opening hours and it can often reduce prices. Similarly, the manufacturer of some widget will often be able to reduce average costs when the scale of production increases. However, if *different* banks join the same ATM network, or if *different* airlines, say, use the same airport, then this can be seen as an example of a market with a platform and two-sided network effects.¹

3 The model

In the following, it will be assumed that cardholders, as well as merchants that accept cards, pay on a per-transaction basis, rather than an annual fee and that non-linear pricing schemes cannot be used. The cardholders pay p_1 per transaction, while merchants pay p_2 . The number of consumers that chooses to hold a card is given by the inverse demand function

$$p_1 = \phi_1(n_1) \tag{1}$$

and the number of merchants that accepts cards is given by inverse demand function

$$p_2 = \phi_2(n_2) \tag{2}$$

where n_1 is the number of cardholders and n_2 is the number of merchants that accept cards, p_i is the per-transaction price for cardholders ($i = 1$) and merchants ($i = 2$), with $\phi' < 0$ for both customer groups. Let $\phi_i(0) = \bar{p}_i$ for $i = 1, 2$ be the maximum price any consumer (merchant) is willing to pay and let $\phi_i(N_i) = 0$, where N_i is the total number of consumers and merchants with non-negative valuations, for $i = 1$ and 2 respectively. Assume that N_1 and N_2 are large numbers and that each identical consumer wishes to make an equal number of transaction with every firm, normalized to 1 for every consumer-merchant pair.

Note that a key modelling assumption is that merchant demand corresponds to social benefits. That is, merchants are not assumed to accept cards for strategic reasons. (Rochet and Tirole, 2004, refer to these modelling assumptions as “Baxter’s case”, after Baxter, 1983.)

Assume that issuing banks have equal and constant marginal costs c_1 per transaction and that acquiring banks’ constant marginal cost is c_2 per transaction. Possibly, there is an interchange fee a , that adds to the marginal cost of the acquiring banks and subtracts from the marginal cost of issuing banks (or

¹Armstrong (2004) provides further examples and analyses network effects in two-sided markets in a general setting.

conversely, for negative values of a). An underlying assumption is that there is just one “platform” (i.e., card system) and that this platform either operates under a not-for-profit basis, or is vertically integrated with issuing and acquisition (i.e., a proprietary system). That is, issues such as double marginalization and access to the bottleneck facilities of a vertically integrated company are assumed away.

The number of merchants that accepts card payments is irrelevant for the decision to become a cardholder and the number of individuals that hold cards is irrelevant for the merchant’s decision to accept card payments. This is so, first, since both groups pay purely on a per-transaction basis (i.e., there are neither per-customer fees, nor fixed customer costs associated with holding a card or maintaining the capacity to accept card payments) and, second, because it is assumed that there are no strategic effects (“Baxter’s case”). That is, it is assumed that merchants do not choose to accept cards in order to induce cardholders to use *their* outlet, rather than another one.²

However, there will be network effects between the two consumer groups. The number of transactions a cardholder makes, and therefore his utility, depends on the number of merchants that accepts cards. Conversely, the number of cardholders affects the number of transactions a merchant makes, as well as her utility. One additional cardholder will increase the consumer surplus of every merchant that accepts cards, while an additional card-accepting merchant will increase the consumer surplus of every cardholder. In other words, although network effects will be irrelevant for the equilibrium outcome in a competitive market (in the absence of subsidy schemes), they appear in the welfare analysis and they will be relevant for parties with market power.

If a fixed annual fee was introduced, then the number of merchants that accepts cards would of course be relevant when a consumer decides whether to adopt a card or not. A consumer i would then become a cardholder only if $(\phi_1(i) - p_1)n_2 \geq F$, where the term within bracket equals consumer i ’s per-transaction surplus, the left-hand side term equals his total surplus and F is the (fixed) annual fee. A corresponding condition would hold for merchants, if they were to pay a fixed annual fee. Although a fixed fee would add to the realism of the model, it will in the following be assumed that only per-transaction fees are used; including fixed fees in the analysis would greatly complicate the analysis.

Given that consumer n_i has adopted a payment card, he will use it with any merchant that accepts cards. Given the assumed (inverse) demand functions above, if consumer n_i uses cards, so will consumers $1, \dots, n_i - 1$; given that merchant n_j accepts cards, so will merchants $1, \dots, n_j - 1$. Under the assumption that each consumer buys once from each merchant, the total number of card transactions will be $n_i n_j$, where n_i and n_j now represent the highest-number consumer and merchant, respectively, that holds a card or accepts card payments. (That is, consumer $n_i + 1$ do not use a payment card and merchant $n_j + 1$ do not accept card payments.)³

²Such a business-stealing effect is analysed by Rochet and Tirole, 2003, and Wright, 2004. See also the discussion in section 11 below.

³The model resembles that of Schmalensee (2002).

4 Perfect competition and welfare maximum

The outcome under perfect competition is straightforward. Since no bank has any market power, prices will be driven to marginal costs. In the absence of an interchange fee, $p_1 = c_1$ and $p_2 = c_2$ will hold. With an interchange fee (issued by a not-for-profit platform), the outcome will be $p_1 = c_1 - a$ and $p_2 = c_2 + a$.

Total welfare will be the sum of the cardholders' and the merchants' valuations of card transactions, less the sum of the costs at the issuing and acquiring sides of the market that these transactions give rise to. That is, welfare is given by

$$W = \int_0^{n_1} (\phi_1(x_1) - c_1) n_2 dx_1 + \int_0^{n_2} (\phi_2(x_2) - c_2) n_1 dx_2 \quad (3)$$

where n_1 and n_2 again represent the number of cardholders and the number of merchants that accept cards, respectively. The first term of the right-hand side integrand - the per-transaction consumer surplus of cardholder x_1 - is multiplied with n_2 , since n_2 merchants accept cards and, therefore, each consumer that holds a card will be able to use the card for n_2 transactions. Similarly, every merchant will meet n_1 cardholders that wishes to make one transaction each.

Evaluating the integral, we find that

$$W = (\Phi_1(n_1) - c_1 n_1) n_2 + (\Phi_2(n_2) - c_2 n_2) n_1 \quad (4)$$

where $\Phi_i(n_i) = \int_0^{n_i} \phi(x_i) dx_i$. Differentiating eq. (4) with respect to n_1 and n_2 , we find the following first-order conditions for welfare maximization

$$\begin{aligned} (\phi_1(n_1) - c_1) n_2 + \Phi_2(n_2) - c_2 n_2 &= 0 \\ (\phi_2(n_2) - c_2) n_1 + \Phi_1(n_1) - c_1 n_1 &= 0 \end{aligned} \quad (5)$$

or, using (1) and (2),

$$\begin{aligned} (p_1 - c_1) n_2 + \Phi_2(n_2) - c_2 n_2 &= 0 \\ (p_2 - c_2) n_1 + \Phi_1(n_1) - c_1 n_1 &= 0 \end{aligned} \quad (6)$$

or

$$\begin{aligned} p_1 &= c_1 - \left(\frac{\Phi_2(n_2)}{n_2} - c_2 \right) \\ p_2 &= c_2 - \left(\frac{\Phi_1(n_1)}{n_1} - c_1 \right) \end{aligned} \quad (7)$$

Since n_2 depends on p_2 and n_1 depends on p_1 , eqs. (7) is a simultaneous-equations system, the solution of which is the pair of optimum prices, p_1 and p_2 . Note that because of the network effects, the solution depends on the global properties of the two demand functions, rather than the marginal properties around the equilibrium (or optimum) point.

In words, the socially optimal cardholder price equals the marginal cost that one transaction gives rise to on the issuing side of the market, minus the difference between the *average* valuation that merchants that accept card payments assigns to a card transaction and the marginal cost of one transaction at the acquiring side. Similarly, the optimal merchant fee equals marginal costs on the acquiring side minus the difference between the cardholders' average valuation and marginal costs on the issuing side. The terms within parenthesis account for the two-sided network effect between cardholders and merchants; the marginal customer on one side of the market will only consider his or her own benefit, not the additional benefit that inframarginal customers on the other side of the market derives from being able to make one additional transaction each.

The following proposition asserts that optimal (linear) prices will result in the card system earning negative profits.

Proposition 1. In welfare-generating card systems, the socially optimal linear transaction prices will be such that the combined revenues of the acquirers and the issuers will be lower than their combined costs.

Proof

The socially optimal prices and the corresponding number of cardholders and merchants that accept cards are defined by eqs. (1), (2) and (7). Let n_1^* and n_2^* be the number of cardholders and card-accepting merchants in optimum. There will be $n_1^*n_2^*$ transactions and each transaction will give an issuer a net profit of $-(\Phi_2(n_2^*)/n_2^* - c_2)$ and an acquirer a net profit of $-(\Phi_1(n_1^*)/n_1^* - c_1)$. Possibly, the net profit will be positive on one of the two sides. However, aggregating over the two sides of the market and multiplying by the number of transactions, total profits will be

$$\begin{aligned} & - \left(\left(\frac{\Phi_2(n_2)}{n_2} - c_2 \right) + \left(\frac{\Phi_1(n_1)}{n_1} - c_1 \right) \right) n_1 n_2 \\ = & -(\Phi_1(n_1) - c_1 n_1) n_2 + (\Phi_2(n_2) - c_2 n_2) n_1 = -W \end{aligned} \quad (8)$$

where the last equality comes from eq. (4). It follows that for any card system that is able to generate positive welfare, the optimal linear transaction prices will not allow issuers and acquirers to recover their combined costs. \diamond

Only if a card system is unable to generate a net surplus should there be no subsidies - but then there should be no card system either.⁴ On the other

⁴If two-part tariffs can be used, efficiency can be improved relative to a situation with only a fixed-fee. For example, if the (inverse) demand curve ϕ does not represent variations in willingness-to-pay between different consumers, but rather an individual (representative) consumer's (cardholder or merchant) willingness-to-pay for additional transactions, then the variable (per-transaction) fee can be set equal to the socially optimal price, while the fixed fee can be set high enough to cover the producers' deficit. In this case, the fixed fees for

hand, if the subsidies are funded by measures that distort allocative efficiency elsewhere, then the welfare gains must be weighed against the welfare losses that are caused by the mechanism used for raising funds.

5 Second-best interchange fees

Given that all cardholders pay the same per-transaction price, p_1 , and similarly for the merchants, given price taking by issuers and acquires and given that the number of transactions is, by definition, the same on both sides of the market, the second-best (or Ramsey) price structure (the optimal price structure in the absence of subsidies) can be achieved with an interchange fee. The first-order condition for welfare maximization is then⁵

$$\frac{dW}{da} = \frac{\partial W}{\partial n_1} \frac{dn_1}{da} + \frac{\partial W}{\partial n_2} \frac{dn_2}{da} = 0 \quad (9)$$

Using (6), the fact that $dn_1/da = -dn_1/dp_1$ and $dn_2/da = dn_2/dp_2$ we have that

$$\begin{aligned} & -((p_1 - c_1)n_2 + \Phi_2(n_2) - c_2n_2) \frac{dn_1}{dp_1} \\ & + ((p_2 - c_2)n_1 + \Phi_1(n_1) - c_1n_1) \frac{dn_2}{dp_2} = 0 \end{aligned} \quad (11)$$

Noting that $p_1 - c_1 = -a$ and $p_2 - c_2 = a$, this expression can be rewritten as

$$(an_2 - \Phi_2(n_2) + c_2n_2) \frac{dn_1}{dp_1} + (an_1 + \Phi_1(n_1) - c_1n_1) \frac{dn_2}{dp_2} = 0 \quad (12)$$

Solving for a gives

$$a = \frac{(\Phi_2(n_2) - c_2n_2) \frac{dn_1}{dp_1} - (\Phi_1(n_1) - c_1n_1) \frac{dn_2}{dp_2}}{n_2 \frac{dn_1}{dp_1} + n_1 \frac{dn_2}{dp_2}} \quad (13)$$

cardholders and merchants, f_1 and f_2 , could be set such that $n_1f_1 + n_2f_2 = W$, leaving the banks with zero profit and the consumers sharing all surplus, also equal to W .

However, if the valuations differ between consumers, a fixed fee will discourage some consumers from adopting cards, even though it would be socially optimal for them to do so. Hence, when determining the optimal fee structure, there will be a trade-off between low marginal fees, so as to encourage cardholders to use the card every time it is efficient to do so, and low fixed fees, so as to encourage low-demand consumers to become cardholders.

For merchants, whether to accept cards or not is an all-or-nothing choice. I.e., a merchant cannot prevent some customers from paying with cards, while allowing others to do so. Hence, given that a merchant has decided to accept cards, a high variable merchant fee will not result in sub-optimal card use in that merchant's facilities. Ceteris paribus, it appears that the merchants' lack of discretion in this respect will lead to lower fixed fees and higher variable fees.

A further analysis of two-part tariffs is beyond the scope of this article. In the following, linear pricing will be assumed.

⁵Cf. proposition 2 in Rochet and Tirole, 2003.

Note that this is not a closed expression, since n_1 and n_2 depends on prices and hence on a . In principle, eq. (13) can be solved for explicit demand functions, but the solutions are likely to be highly nonlinear. However, multiplying both the numerator and denominator on the right-hand side with $p_1 p_2 / n_1 n_2$ gives another expression, which can provide some intuition

$$a = \frac{p_2 \left(\frac{\Phi_2(n_2)}{n_2} - c_2 \right) \varepsilon_1 - p_1 \left(\frac{\Phi_1(n_1)}{n_1} - c_1 \right) \varepsilon_2}{p_2 \varepsilon_1 + p_1 \varepsilon_2} \quad (14)$$

where $\varepsilon_i = \frac{\partial n_i}{\partial p_i} \frac{p_i}{n_i}$.

According to eq. (14), subsidies will tend to flow towards the issuing side of the market if cardholders' demand is more elastic than merchants' demand and if the difference between the average valuation of the merchants and the acquirers' cost in equilibrium is high relative to the difference between the average valuation of the cardholders and the issuers' cost - also evaluated in equilibrium.

Under the assumptions of perfect competition on both sides of the market and in the context of the present model, there is no point trying to use the interchange fee as a collusion device: issuers and acquirers will still make zero profit. In fact, as noted by Schmalensee (2002), if both acquires and issuers are perfectly competitive, there are no reasons for them to have any particular preferences over the interchange fee at all.

6 Bilateral monopolies

In this section, it will be assumed that the market is controlled by two monopolies: one monopoly issuer and one monopoly acquirer. This is the simplest way to introduce market power in the model except, possibly, for assuming a single monopoly provider of both issuing and acquiring services. The latter alternative, however, also implies that a single entity can internalize network effects between the two sides of the market.

If the two monopolies take a as given, they will maximize the following profit functions

$$\begin{aligned} \pi_1 &= [\phi_1(n_1) - c_1 + a] n_1 n_2 \\ \pi_2 &= [\phi_2(n_2) - c_2 - a] n_2 n_1 \end{aligned} \quad (15)$$

by choosing n_1 (the issuer) and n_2 (the acquirer). The monopoly issuer will take n_2 as given, while the monopoly acquirer will take n_1 as given. Using $\phi_1(n_1) = p_1$ and $\phi_2(n_2) = p_2$, the first-order conditions will be

$$\begin{aligned} p_1 &= c_1 - a - \phi_1'(n_1) n_1 \\ p_2 &= c_2 + a - \phi_2'(n_2) n_2 \end{aligned} \quad (16)$$

Hence, the two monopolies will price above their marginal costs, in contrast to the below-cost pricing required for achieving the social optimum. As noted by

Schmalensee (2002), the interchange fee does not resolve the problem of double marginalization, since it only shifts marginal costs from one side of the market to the other. However, it can mitigate problems that are due to differences (in terms of elasticities and average valuations of inframarginal customers) between cardholders and merchants.⁶

7 A proprietary system (two-sided monopoly)

A two-sided monopoly - e.g., a proprietary system, such as American Express, although in a monopoly position - would maximize the sum of its profits generated on the two sides of the market, i.e.

$$\pi = [\phi_1(n_1) - c_1]n_1n_2 + [\phi_2(n_2) - c_2]n_1n_2 \quad (17)$$

Obviously, an interchange fee would play no role for a proprietary system. The first-order conditions for profit maximization, obtained by differentiating eq. (17) with respect to n_1 and n_2 , will be

$$\begin{aligned} [\phi'_1(n_1)n_1 + \phi_1(n_1) - c_1] + [\phi_2(n_2) - c_2] &= 0 \\ [\phi'_2(n_2)n_2 + \phi_2(n_2) - c_2] + [\phi_1(n_1) - c_1] &= 0 \end{aligned} \quad (18)$$

Substituting $p_1 = \phi_1(n_1)$ and $p_2 = \phi_2(n_2)$ into the above equations, we have

$$\begin{aligned} \phi'_1(n_1)n_1 + p_1 - c_1 + p_2 - c_2 &= 0 \\ \phi'_2(n_2)n_2 + p_2 - c_2 + p_1 - c_1 &= 0 \end{aligned} \quad (19)$$

Or, alternatively,

$$p_1 + p_2 - c_1 - c_2 = -\phi'_1(n_1)n_1 = -\phi'_2(n_2)n_2 \quad (20)$$

The sum of the mark-up above costs on the two sides of the market should, hence, be equal to the revenue increase on inframarginal units on *one* side of the market, if output is reduced by one unit on that side.⁷ Alternatively, one could say that when the monopolist considers increasing production by one unit on one side of the market, its marginal (per-transaction) costs will be $c_1 + c_2$, since each transaction necessarily involves both the issuing and the acquiring side. At the same time, its marginal revenues will include revenues generated from the additional sales on both sides of the market, *but only losses from price reductions on inframarginal units on one side of the market*. For example, the monopoly may consider lowering the price on the issuing side of the market. This will generate additional sales on the issuing side of the market, but also on the acquiring side, since each transaction involves both issuing and acquiring. Additional sales implies additional costs and revenues on both sides of the market, but the price reduction will reduce its revenues from inframarginal

⁶Schmalensee provides further results on interchange fees under bilateral monopoly.

⁷See also Proposition 1 in Rochet and Tirole, 2003.

customers on the issuing side of the market only. Naturally, the monopoly should optimize against both sides of the market, so that the revenue increase on inframarginal customers that would result if the number of customers were reduced should be equal on the two sides of the market.

Adding the respective first-order conditions of the two sellers in a bilateral monopoly, eqs. (16), we find that $p_1 + p_2 - c_1 - c_2 = -(\phi'_1(n_1)n_1 + \phi'_2(n_2)n_2)$. In other words, for given values of n_1 and n_2 the combined mark-up in a bilateral monopoly will be exactly twice as large as that in a proprietary system. However, n_1 and n_2 will in general *not* be the same under the two market configurations. For example, if ϕ'_i is constant for $i = 1, 2$ and if prices are lower under a proprietary system, then n_i will be higher. Consequently, the terms $-\phi'_i(n_i)n_i$ will increase relative to the bilateral monopoly situation, implying that the sum of the price-cost margins under a proprietary system will be greater than one half of the sum of the price-cost margins under a bilateral monopoly.

Proposition 2. With linear demand curves, the combined price-cost margins of the two sides under a proprietary system will equal 2/3 of the combined price-cost margins under bilateral monopoly.

Proof.

Let the inverse demand functions ϕ_i be linear, such that $p_i = \phi_i(n_i) = a_i - b_i n_i$. Straight-forward calculations will show that prices under a proprietary system will be given by $p_1 = (2a_1 - b_2 + c_1 + c_2)/3$ and $p_2 = (2a_2 - a_1 + c_1 + c_2)/3$, while under bilateral monopoly the prices will be $p_1 = (a_1 + c_1 - \mathbf{a})/2$ and $p_2 = (a_2 + c_2 + \mathbf{a})/2$, with the interchange fee, \mathbf{a} , written in bold to avoid confusion. Hence, the combined price-cost margin under a proprietary system will be

$$p_1 + p_2 - c_1 - c_2 = \frac{a_1 + a_2 - c_1 - c_2}{3}$$

while the combined price-cost margin under bilateral monopoly will be

$$p_1 + p_2 - c_1 - c_2 = \frac{a_1 + a_2 - c_1 - c_2}{2}$$

The former price-cost margin is 2/3 of the latter. \diamond

It is evident that total profits will be at least as high under a proprietary system as under bilateral monopoly. A proprietary system can replicate the pricing structure of a bilateral monopoly, but it can also avoid double marginalization.

Furthermore, it seems likely that a proprietary system will set lower prices than a bilateral monopoly. Assume that the inverse demand functions ϕ_i are as given in Proposition 2. Using the first-order conditions for profit maximization under the two market configurations, eqs. (19) and (16), and solving for n_i , we

get

$$\begin{aligned}
n_1^{prop} &= \frac{a_1 + a_2 - c_1 - c_2}{3b_1} \\
n_2^{prop} &= \frac{a_1 + a_2 - c_1 - c_2}{3b_2} \\
n_1^{bil} &= \frac{a_1 - c_1 + \mathbf{a}}{2b_1} \\
n_2^{bil} &= \frac{a_2 - c_2 - \mathbf{a}}{2b_2}
\end{aligned} \tag{21}$$

where superscript *prop* and *bil* represent the proprietary system and the bilateral monopoly, respectively. It is clear that if the two sides of the markets are symmetric, i.e., if $a_1 = a_2$ and $b_1 = b_2$, then \mathbf{a} is optimally set equal to 0 and the number of cardholders and merchants will be $n_1^{prop} = n_2^{prop} = \frac{2(a_1 - c_1)}{3b_1}$ and $n_1^{bil} = n_2^{bil} = \frac{a_1 - c_1}{2b_1}$. That is, the number of agents that adopt cards will be 50 per cent larger on each side under a proprietary system. (As shown above, the price-cost margin will be 50 per cent higher under bilateral monopoly.) The number of transactions, given by $n_1 n_2$, will be 9/4 times as high under a proprietary system.

If the two markets are not symmetric, while the inverse demand functions are still linear, from eq. (3), welfare will be given by

$$W = n_1 n_2 (a_1 - \frac{1}{2} b_1 n_1 - c_1 + a_2 - \frac{1}{2} b_2 n_2 - c_2) \tag{22}$$

Proposition 3. For general linear inverse demand functions, the number of cardholders, card-accepting merchants and transactions will be higher under a proprietary system than under bilateral monopoly, if the interchange fee is set so as to maximize the sum of the acquirer and issuer profits. In additional, welfare is higher under a proprietary system.

Proof.

Eq. (21) gives the quantities that will be chosen in a proprietary system, as well as the quantities that acquirers and issuers will chose, given the interchange fee a . The combined profit of the issuers and the acquirers is

$$\begin{aligned}
\pi &= \pi_1 + \pi_2 = \\
&= [a_1 + b_1 n_1 - c_1 + a] n_1 n_2 + [a_2 + b_2 n_2 - c_2 - a] n_1 n_2 = \\
&= \frac{1}{8b_1 b_2} [a_1 + a_2 - c_1 - c_2] [a_1 - c_1 + a] [a_2 - c_2 - a]
\end{aligned} \tag{23}$$

Using the first-order condition with respect to a , it follows that $a = (a_2 - a_1 - c_2 + c_1)/2$. Inserting this in eq. (21) gives

$$\begin{aligned} n_1^{bil} &= \frac{a_1 - c_1 + a_2 - c_2}{4b_1} \\ n_2^{bil} &= \frac{a_1 - c_1 + a_2 - c_2}{4b_2} \end{aligned}$$

Comparing these quantities for the case of bilateral monopoly with the quantities chosen by a proprietary system, as given by eq. (21) proves the proposition's statements on quantities. Since quantities are sub-optimal, it follows that welfare increases with quantities. \diamond

8 One-sided monopoly

An alternative assumption is that there is market power on *one* of the two sides of the market - e.g., on the issuing side - but not on the other. I assume that the acquiring side of the market is competitive, while the issuing side is controlled by a monopolist⁸, but that the interchange fee is again restricted to be zero. Now the issuer will want to maximize the following profit function:

$$\pi_1 = [\phi_1(n_1) - c_1]n_1n_2 \quad (24)$$

As under bilateral monopoly, the issuing monopoly will consider the number of merchants, n_2 , as fixed (as long as an interchange fee is not available; see below). Hence, using $p_1 = \phi_1(n_1)$, the first-order condition for profit maximization is

$$p_1 = c_1 - \phi_1'(n_1)n_1 \quad (25)$$

Note that the first-order condition is identical to that of the issuer under bilateral monopoly when a is restricted to be 0. It follows that welfare is higher than under bilateral monopoly (and zero interchange fee), but lower than under competition (and no interchange fee).

Note also that the mark-up on the issuer side now equates revenue loss on inframarginal units from increasing sales with one unit, while in a proprietary system the sum of the mark-ups on the two sides of the market were optimally set equal to the loss on inframarginal units from increasing sales one unit (on one side of the market).

If the issuer holds monopoly power vis-à-vis the cardholders, while the acquiring side is competitive, there appears to be good reasons to think that the issuer holds monopoly power also with respect to the acquirers. If this is the

⁸Rochet and Tirole (2003) argue that the acquiring side is more competitive than the issuing side. In contrast to the present paper, they assume that the issuing side is characterized by a symmetric oligopoly. My assumption of a monopolized issuing market could perhaps be justified on the grounds that bank customers - at least non-business debit-card customers - are captive. I.e., bank customers will not switch bank in order to get a better price on card transactions.

case, it can set an interchange fee that will determine the price on the acquiring side of the market; given that the acquirers are competitive, they will set $p_2 = c_2 + a$. Since the interchange fee is paid to a monopolist, however, it will have no effect on the issuing side of the market. Therefore, the one-sided monopoly will, arguably, have the same power to set two prices as a proprietary monopolist has (unless multilateral agreements face stronger regulatory resistance). Accordingly, the issuer will maximize eq. (17) and the price structure will be identical to that of the proprietary system.

9 Welfare comparisons

So far, I have analysed price-setting under six different regimes: perfect competition without an interchange fee, perfect competition with a socially optimal interchange fee, the first-best solution, bilateral monopoly (with and without an interchange fees), one-sided monopoly (without an interchange fee) and a single monopoly/proprietary system. As argued above, a one-sided monopoly which can also determine the interchange fee will be able to replicate the outcome chosen by the proprietary monopoly. Consequently, there is no need to analyse this case separately.

Proposition 4. The welfare effects of the ownership structure of the card industry can be characterized as follows:

- a) Welfare under socially optimal prices is higher than welfare under competition, also for a socially optimal interchange fee.
- b) Welfare under competition and with an optimal interchange fee is higher than under either of the following: a proprietary system or competition without an interchange fee.
- c) Welfare under competition and without an interchange fee could be either higher or lower than that under a proprietary system.
- d) Welfare is lower under bilateral monopoly, where the two firms set the interchange fee so as to maximize combined profits on the two sides of the market, than under a proprietary system.
- e) In the absence of interchange fees, welfare is higher under a one-sided monopoly than under bilateral monopoly, and higher still under competition.
- f) Given the existence of a monopoly on one side of the market and perfect competition on the other side, introducing an access fee can either increase or decrease welfare.

Proof.

- a) From eq. (7) we know that the optimal issuing price is $p_1^{opt} = c_1 - (\phi_2(n_2)/n_2 - c_2)$. Under competition, the issuing price will be $p_1 = c_1 - a$. Using eq. (13), the issuing price with a socially optimal

interchange fee will be

$$p_1^{int} = c_1 - \omega \left(\frac{\phi_2(n_2)}{n_2} - c_2 \right) + (1 - \omega) \left(\frac{\phi_1(n_1)}{n_1} - c_1 \right) \quad (26)$$

where

$$\begin{aligned} \omega &= \frac{n_2 \frac{dn_1}{dp_1}}{n_2 \frac{dn_1}{dp_1} + n_1 \frac{dn_2}{dp_2}} \\ 1 - \omega &= \frac{n_1 \frac{dn_2}{dp_2}}{n_2 \frac{dn_1}{dp_1} + n_1 \frac{dn_2}{dp_2}} \end{aligned} \quad (27)$$

Given that the card system generates welfare, the expressions $(\phi_2(n_2)/n_2 - c_2)$ and $(\phi_1(n_1)/n_1 - c_1)$ are both positive. From eq. (27) it is obvious that ω and $(1 - \omega)$ are positive. Let the values of n_1 and n_2 corresponding to the prices p_1^{opt} and p_1^{int} be given by p_i^{opt} and p_i^{int} , respectively. Assuming that $\phi_i(n_i) < 0$ and subtracting p_1^{opt} from p_1^{int} yields

$$p_1^{int} - p_1^{opt} = (1 - \omega) \left[\left(\frac{\phi_1(n_1^{int})}{n_1^{int}} - c_1 \right) + \left(\frac{\phi_2(n_2^{int})}{n_2^{int}} - c_2 \right) \right] + \left[\frac{\phi_2(n_2^{opt})}{n_2^{opt}} - \frac{\phi_2(n_2^{int})}{n_2^{int}} \right]$$

The first term on the right-hand side is positive, while the second term is negative if $p_1^{int} - p_1^{opt} > 0$ and positive if $p_1^{int} - p_1^{opt} < 0$. Assume that $p_1^{int} - p_1^{opt} < 0$. Then both terms on the right-hand side are positive. This is a contradiction; it follows that $p_1^{int} - p_1^{opt} > 0$. Since the price under competition differs from the optimal price, it follows that welfare is lower.

b) Comparing a proprietary system with a competitive system with an optimal interchange fee, we know that in the former, $p_1 + p_2 = c_1 + c_2 - \phi_1'(n_1)n_1 = c_1 + c_2 - \phi_2'(n_2)n_2$, while in the latter $p_1 + p_2 = c_1 + c_2$. That is, the total price-cost margin will be higher in a proprietary system. Since the optimal interchange fee is set so that welfare is maximized given that p_1 and p_2 sum to $c_1 + c_2$, welfare cannot be higher when p_1 and p_2 sum to something larger than that. It follows that welfare is higher under an optimal interchange fee than under a proprietary system (or one-sided monopoly with competition on the other side, where the monopoly can use an interchange fee).

Comparing a competitive system with or without an (optimally set) interchange fee, it is obvious that a system without an interchange fee will result in lower welfare, except when the optimal interchange fee is zero.

c) From b) and from the fact that the optimal interchange fee can be zero, it follows that welfare can be higher under competition without

an interchange fee than for a proprietary system. To demonstrate the possibility that the opposite claim can be true, it is sufficient to consider a market where the marginal valuation of cardholders is constant and ε below c_1 , where ε is a value close to zero, while the marginal valuation of merchants is downward sloping and higher than c_2 for some merchants. In the absence of an interchange fee and under competition, there will be no card system, while a proprietary system will find it profitable to subsidize cardholders with ε per transaction.

d) Assume that $\phi'_i(n_i)n_i$ is decreasing in n_i . (Obviously, this is true for linear demand functions, since $\phi'_i(n_i)$ is then constant and negative.) From the first-order condition for profit maximization in bilateral monopoly, eq. (16), we know that $\phi'_1(n_1^{bil})n_1^{bil} + p_1^{bil} - c_1 + a = 0$ (and correspondingly for firm 2), where n_i^{bil} and p_i^{bil} are the quantities and prices, respectively, that will result in a bilateral monopoly where a is set so as to maximize combined profits. From a proprietary firm's (a two-sided monopoly's) first-order condition for profit maximization, eq. (19), we know that $\phi'_1(n_1^{pro})n_1^{pro} + p_1^{pro} - c_1 + p_2^{pro} - c_2 = 0$. The latter equation can be rewritten $\phi'_1(n_1^{pro})n_1^{pro} + p_1^{pro} - c_1 + a + p_2^{pro} - c_2 - a = 0$, where a is the interchange fee that maximizes total profits under a bilateral monopoly.

We know that in the bilateral monopoly quantities will be chosen so that $p_2^{bil} - c_2 - a > 0$ and $p_1^{bil} - c_1 + a > 0$. Otherwise the firms would not be active in the market. It follows that $\phi'_1(n_1^{bil})n_1^{bil} + p_1^{bil} - c_1 + a + p_2^{bil} - c_2 - a > 0$. Comparing the two first-order conditions and noting that both $p_1(n_1)$ and $\phi'_1(n_1)n_1$ are decreasing in n_1 , while c_1 and c_2 are fixed, it follows that n_1^{pro} must be greater than n_1^{bil} . A similar analysis shows that n_2^{pro} must be greater than n_2^{bil} . Since welfare increases in n_i for n_i smaller than the optimal quantities, it follows that welfare is higher for a proprietary firm than under bilateral monopoly.

e) Under competition, the first-order conditions that determine prices and quantities will be $p_1 = c_1$ and $p_2 = c_2$. From eqs. (25) and (16) the corresponding first-order conditions for one-sided and bilateral monopolies will be $p_1 = c_1 - \phi'_1(n_1)n_1$ for the issuing side and $p_2 = c_2$ and $p_2 = c_2 - \phi'_2(n_2)n_2$, respectively, for the acquiring side. From eq. (7) we know that the optimal price is lower than c_i , but since $\phi'_i(n_i) < 0$ we see that monopolies will set prices higher than c_i . It follows that welfare is lower under a one-sided monopoly than under competition, and lower still in a bilateral monopoly.

f) Using the linear demand system representations $p_1 = a_1 - b_1n_1$ and $p_2 = a_2 - b_2n_2$ it can be shown that welfare under a one sided

monopoly without an interchange fee will be

$$W_{1-sided} = \frac{1}{8b_1b_2} (a_1 - c_1)(a_2 - c_2)(3a_1 + 2a_2 + c_1 - 2c_2) \quad (28)$$

while welfare under a proprietary system will be

$$W_{prop} = \frac{2}{27b_1b_2} (a_1 + a_2 - c_1 - c_2)^3 \quad (29)$$

It can easily be demonstrated that neither of the two regimes dominates the other. For example, with parameter values $a_1 = a_2 = 10$, $b_1 = b_2 = 1$ and $c_1 = c_2 = 2$, a one-sided monopoly is preferable to a proprietary system. On the other hand, if $a_2 = 3$ and the other parameters are unchanged, welfare will be higher under a proprietary system. \diamond

According to the proposition, there is no unique ranking between a proprietary system and competition (without an interchange fee). If the two sides of the markets are perfectly symmetric - so that the optimal level of the interchange fee is $a = 0$ - then a competitive market will perform just as well as a market with an optimal interchange fee and competition. It follows that a competitive market (without an interchange fee) outperforms a proprietary system in a symmetric setting.

On the other hand, if the two sides of the market are highly asymmetric, then welfare will be higher if one side of the market subsidizes the other. That means that a proprietary system (with an explicit or implicit interchange fee) may be preferable to a competitive system (without an interchange fee). One example would be a situation where demand on the issuing side of the market, say, is such that all potential cardholders have a valuation that is marginally below the issuers' cost, while merchants have a relatively inelastic demand for card services, such that their average valuation is much higher than acquirers' marginal costs. Under perfect competition, price would equal marginal costs and, hence, no one would become a cardholder. However, it would be in a monopolist's interest to set the interchange fee such that most or all potential cardholders would find it in their interest to become actual cardholders. This would generate a large surplus on the *acquiring* side of the market. Even though a monopolist would be able to extract a relatively large share of that surplus, welfare would still be higher than if there were no card transactions at all.

In contrast, if an interchange fee is not available, there *is* a unique ranking that puts competition ahead of a one-sided monopoly, and the latter ahead of a bilateral monopoly. Without an interchange fee, firms with market power cannot achieve subsidization between the two sides of the market, but they will obtain a positive mark-up.

10 Oligopoly markets

In the previous sections, the polar cases of monopoly and perfect competition have been analysed. Although these cases provide some basic intuition on the functioning of two-way network markets, it may also be desirable to allow for intermediate levels of market power. There appears to be at least three segments of the market where an assumption of either perfect competition or monopoly may be too simplistic. First, there can be oligopoly interaction between banks, either between issuers or between acquirers (or both). Second, the merchants can have market power vis-à-vis their customers. Finally, more than one card system may compete for customers - so-called platform competition. Literature relating to these three alternatives will be briefly surveyed below.

Merchant market power

In the literature that analyses the effect of merchant market power, a typical assumption is that merchants have market power (a positive price-cost margin) in relation to their customers, while being price takers in relation to acquiring banks. As mentioned in the introduction, the result may be overprovision of payment cards. The reason is that the merchants' *private* valuation of card services will exceed the social value and that acquirers will be able to extract part of that value, which may then be used to induce (excessive) card adoption on the issuing side.

More specifically, the above mechanism will be effective if at least some consumers are aware of which merchants that accept cards and are influenced in their choice of merchant by this aspect of service quality. Alternatively, some consumers may search until they find a merchant that accepts cards. Then, a merchant will have a *strategic* interest in accepting card payments, in addition to the intrinsic benefits of receiving card payments rather than cash. Accepting card payments will increase the merchant's sales, at the expense of other merchants. This strategic interest can be exploited by acquiring banks with market power, or by issuing banks through an interchange fee, as shown by Rochet and Tirole (2002). The acquirers will be able to charge a higher price than otherwise. However, the gains the merchants will be able to make at the expense of one another will not correspond to net welfare gains. The additional surplus the acquirers are able to extract will, therefore, lead to higher consumer prices. In effect, assuming that the merchants cannot price discriminate between card customers and non-card customers, the latter category of consumers will be subsidizing the former.⁹ The welfare effect of this subsidy, however, can be either positive or negative. Because of network effects, it may be welfare-increasing to subsidize cardholders, but the subsidies may also be excessive.

Rochet and Tirole (2002), Gans and King (2003) and Wright (2004) all find that under certain conditions, the profit-maximizing interchange fee may deviate from the optimal one, while under other conditions they may coincide. (For a

⁹Rochet and Tirole derive this result under the assumption of price competition in differentiated products. Under the assumption of price competition in homogenous products, non-card customers would go to merchants that do not accept cards, while card customers would patronize merchants that do.

discussion of these results, see Chakravorti, 2003, Section 3.2.)

Platform competition

Although a discussion of platform competition falls outside the scope of this paper, a few references will be provided. Rochet and Tirole (2003) provide a general framework for analysing platform competition. Guthrie and Wright (2003) find that competition between platforms is ineffective, unless cardholders hold just one card, while Chakravorti and Roson (2004) find that competition unambiguously increases consumer and merchant welfare.

Oligopolistic banks

The analysis becomes less tractable under oligopoly interaction. An intuitively plausible result is that the interchange fee will be used to transfer profit from the more competitive side of the market to the less competitive side. This is formally shown by Manenti and Somma (2003), in a setting with one not-for-profit platform and one proprietary system. They assume that the not-for-profit platform maximizes the sum of issuers' and acquirers' profits and that competition between banks is such that there will be a fixed mark-up above issuers' and acquirers' marginal costs. For example, given issuers' and acquirers marginal costs c_1 and c_2 , prices will be αc_1 on the issuing side and βc_2 on the acquiring side, where $\alpha, \beta > 1$. With an interchange fee a , equilibrium prices will instead be $\alpha(c_1 - a)$ and $\beta(c_2 + a)$. A more surprising result is that the banks' total profit will not depend on α or β . In particular, if competition increases on one side of the market, i.e., if α or β falls, then the interchange fee will be rebalanced so that total profit remains unchanged.

Assuming fixed mark-ups can be seen as a reduced-form analysis. In the following, the behaviour of the oligopoly firms will instead be explicitly modelled. In order to simplify the analysis, the acquiring side of the market is assumed to be competitive and the merchants are assumed to have no strategic reasons for accepting cards (the "Baxter case"). Except that there are two issuers that compete in quantities, the set-up is the same as in the previous sections. Demand is given by eqs. (1) and (2), but with the sum of the two firms' quantities as the argument in ϕ_1 . The two issuers are assumed to have identical and constant marginal cost, c_1 . It follows that the profit function of the issuers will be

$$\pi_{1i} = [\phi_1(n_{11} + n_{12}) - c_1 + a]n_{1i}n_2 \quad (30)$$

where n_{1i} now denotes the number of card customers that bank $i = 1, 2$ has, such that $n_1 = n_{11} + n_{12}$.¹⁰ The corresponding first-order condition for issuer i is

$$\phi_1'(n_{11} + n_{12})n_{1i} + \phi_1(n_{11} + n_{12}) - c_1 + a = 0 \quad (31)$$

Exploiting the fact that the two issuers are symmetric, we have that

$$\phi_1'(2n_{11})n_{11} + \phi_1(2n_{11}) - c_1 + a = 0 \quad (32)$$

¹⁰Note that subindex 1 can now, depending on the circumstances, refer either to market 1 (the issuing market) or firm 1. Hence, while π_1 denotes the profit of firm 1, p_1 denotes prices in the issuing market.

In order to proceed, a more specific model will be used in the following. Assume now that demand is linear and given by inverse demand function

$$p_1 = L - n_{11} - n_{12} \quad (33)$$

on the issuing side of the market, while (inverse) demand on the competitive acquiring side of the market is given by

$$p_2 = M - n_2 \quad (34)$$

where n_2 , as before, is the total number of customers (merchants) of all acquiring banks. As before, the marginal costs are denoted c_1 and c_2 . It is straight-forward to show that the equilibrium quantities, for a given level of a , will be $n_{11} = n_{12} = (L - c_1 + a)/3$, which will result in $p_1 = L/3 + 2(c_1 - a)/3$. This, in turn, will yield profits of

$$\pi_1 = \pi_2 = ((L - c_1 + a)/3)^2 n_2 = ((L - c_1 + a)/3)^2 (M - c_2 - a) \quad (35)$$

where the last equality comes from the merchants' inverse demand function and from the assumption of perfect competition between the acquirers. For the particular case of $a = 0$, prices will be $p_1 = \frac{1}{3}(L + 2c_1)$ and $p_2 = c_2$.

If the duopolists can jointly set the interchange fee, they will set it so as to maximize the above expression, i.e., such that

$$a = \frac{2}{3}M - \frac{1}{3}L + \frac{1}{3}c_1 - \frac{2}{3}c_2 = \frac{1}{3}(2M - L - c_1 - 2c_2) \quad (36)$$

This will result in prices

$$p_1 = L/3 + 2(c_1 - \left(\frac{2}{3}M - \frac{1}{3}L + \frac{1}{3}c_1 - \frac{2}{3}c_2\right))/3 = \quad (37)$$

$$= \frac{5}{9}L - \frac{4}{9}M + \frac{4}{9}c_1 + \frac{4}{9}c_2 \quad (38)$$

$$\begin{aligned} p_2 &= c_2 + \frac{2}{3}M - \frac{1}{3}L + \frac{1}{3}c_1 - \frac{2}{3}c_2 = \\ &= \frac{2}{3}M - \frac{1}{3}L + \frac{1}{3}c_1 + \frac{1}{3}c_2 \end{aligned} \quad (39)$$

and quantities

$$n_1 = 2n_{11} = 2n_{12} = \frac{4}{9}(L + M - c_1 - c_2) \quad (40)$$

$$n_2 = \frac{1}{3}(L + M - c_1 - c_2)$$

To simplify further, assume that $M = L$ and that $c_1 = c_2$. Then

$$\begin{aligned} a &= \frac{1}{3}(L - c_1) \\ p_1 &= \frac{1}{9}(L + 8c_1) \\ p_2 &= \frac{1}{3}(L + 2c_1) \end{aligned} \quad (41)$$

This means that the two-firm Cournot price will be set in the *acquiring* market, while price in the issuing market will be identical to the *eight*-firm Cournot price.

In the general (linear) case, corresponding to prices given by eq. (37) and interchange fee given by eq. (36), prices will fall in the issuing market and rise in the acquiring market if $(L - c_1) < 2(M - c_2)$, while if $(L - c_1) > 2(M - c_2)$ prices will *rise* in the issuing market and *fall* in the acquiring market (relative to a situation with $a = 0$). The interpretation is that if approximately equal amounts of consumer surplus are generated in both markets, the duopoly issuers will set the interchange fee so as to extract profit from the acquiring market. This will lower the marginal cost in the issuing market. Hence, as a side-effect, prices will fall on that market. However, if much more consumer surplus is generated in the issuing market, then the duopolists should in fact set the interchange fee so as to *subsidize* the acquiring market (and tax themselves). This will reduce the merchant discount and increase the number of merchants. This, in turn, increases the number of transactions per cardholder, with the ultimate effect of increasing the issuers' revenues.

The welfare effect of the interchange fee can be evaluated by comparing welfare when the interchange fee is set as prescribed by eq. (36) and welfare when the interchange fee is set to zero, as the following proposition shows.

Proposition 5. Welfare under a one-sided duopoly where the firms compete in quantities, where there is competition on the other side and where the duopolists can set the interchange fee so as to maximize their combined profits can be either higher or lower than welfare under the same market structure with the interchange fee restricted to zero.

Proof

Simple calculations show that when the interchange is fixed at zero, quantities will be given by

$$\begin{aligned} n_1 &= \frac{2}{3}(L - c_1) \\ n_2 &= M - c_2 \end{aligned} \tag{42}$$

i.e., Cournot quantities in the issuing market and the competitive quantity in the acquiring market. Welfare for the two cases can be calculated using eq. (22), but with L taking the place of a_1 , M replacing a_2 and $b_1 = b_2 = 1$. Inserting quantities, given by eqs. (40) and (42), respectively, gives

$$\begin{aligned} W^{IF} &= \frac{22}{243}(L + M - c_1 - c_2)^3 \\ W^0 &= \frac{1}{9}(L - c_1)(M - c_2)(4L + 3M - 4c_1 - 3c_2) \end{aligned} \tag{43}$$

where W^{IF} denotes welfare when the duopolists set a profit-maximizing interchange fee, according to eq. (36) and W^0 denotes welfare when there is no (or zero) interchange fee. It is easily shown that neither of the two cases dominates the other, for example by using the numerical example from Section 9. \diamond

In line with intuition, when the two sides of the markets are symmetric, welfare is higher if the interchange fee is fixed at zero, while if demand is much smaller on the acquiring side, a profit-maximizing interchange fee will increase welfare.

11 Introducing system costs

Above, it has implicitly been assumed that there are no system costs. The only costs that arise are the issuers' and the acquirers' marginal costs. However, it is likely that there will be processing costs in the central system, which the system will have to cover by charging acquirers and/or issuers. Schmalensee (2002) notes that if there are no restrictions on how the transaction costs are allocated between issuers and acquirers, then an interchange fee may be redundant. Instead of a positive interchange from, say, acquirers to issuers, a larger fraction of the system costs could be allocated to the acquirers. Schmalensee suggests that due to transactions costs, such reallocations may not always be feasible. In addition, a non-zero interchange fee may have a role to play in corner solutions. Even if all system costs are allocated to one side of the market, effective marginal cost may still be above the optimal level on the other side of the market.

It is likely that some of the system costs will be variable and some fixed. The system (or the banks) can decide that these costs should be allocated to the participating banks, either as fixed fees or as variable fees (or as a combination of both). In principle, variable costs could be included in a fixed fee, for example by setting a high annual fee that gives a participating bank the right to make any number of transactions through the system, free of charge. In practice, if total variable costs are substantial, it appears reasonable to assume that these will be allocated to the participating banks as variable fees. Otherwise, it would be apparent that some banks were subsidizing others.

From an analytical point-of-view, the above argument suggests that system costs that are variable with the number of transactions can be seen as part of the issuers' and acquirers' variable costs. Consequently, there is no point in complicating the analysis by introducing variable system costs.

Fixed system costs, on the other hand, give rise to genuinely new issues. The collective of banks can decide that fixed system costs should be covered with variable fees, or with fixed fees, and this choice is likely to have a material effect. Allocating fixed costs in proportion to the number of transactions will, from an individual bank's point-of-view, transform fixed costs into variable costs. This will raise marginal costs which, in turn, will tend to make the downstream

issuing and acquiring markets less competitive. On the other hand, high fixed fees will serve as an entry barrier for small banks and, as argued by Hausman *et al.* (2003), will reduce the banks' incentives to invest in technologies that reduces marginal costs. The intuition behind the latter argument is that if competition between banks is strong, downstream prices will be close to marginal costs and fixed cost (e.g., due to technology investments) will not affect the equilibrium price. If the banks' marginal costs are reduced because of investments, prices will fall in proportion, but the investment costs will have to be met by increased fixed fees. Consequently, the banks will bear the investment costs, while the consumers will be reaping the benefits.

It appears that fixed costs in practice often *are* covered with variable fees. One reason may be that it is difficult to get acceptance for an equal division of fixed costs between banks of different sizes - or indeed for any other division in fixed proportions - while distributing fixed costs in proportion to the number of transactions is likely to be perceived as fair.

To begin an analysis of the effect of fixed costs in the central system, let $n_1 n_2$ represent the number of transactions, s the number of issuers and B the fixed system costs. Assume, for simplicity, that B is allocated between issuers only. If the fixed costs are divided in proportion to the number of transactions, then issuer i 's per-transaction marginal cost will be $c_1 - a + B/(n_1 n_2)$ and its total cost will be $n_{1i} n_2 (c_1 - a + B/(n_1 n_2))$, where n_{1i} is the number of card customers i has and $n_{1i} n_2$ is the number of transactions processed by issuer i . If B is instead divided equally between all issuers, irrespective of how many transactions they make, then marginal and total costs will be $c_1 - a$ and $n_{1i} n_2 (c_1 - a) + B/s$, respectively.

Under perfect competition, the issuers and the acquires will set prices equal to marginal costs. As argued above, they will be indifferent as to the level of the interchange fee, since they will make zero profit irrespective of its level. However, marginal-cost pricing in combination with a fixed-proportion allocation of B between the banks would result in negative profits. In a static perspective, this would be welfare improving, but the situation would be unsustainable. For a long-run analysis, there are two natural modelling approaches. One is to make entry endogenous, so as to ensure that the banks' profits will be non-negative, i.e., a free-entry equilibrium (Section 13). The other is to assume that issuers are oligopolistic (Section 14).

To recapitulate, the trade-offs will be as follows. If fixed system costs are allocated in proportion to the number of transactions, the banks will experience higher marginal costs. This will raise prices, causing them to depart (further) from the first-best optimum. On the other hand, entry barriers will be low. This will stimulate entry which, in turn, will exert a downward pressure on prices. Furthermore, following the analysis of Hausman *et al.*, the banks' resistance to investing in technology that reduces marginal cost will not be strong.

If, instead, fixed costs are allocated in fixed proportions between the banks, prices will tend to be lower because of the lower marginal costs. On the other hand, there will be less entry, which will tend to increase prices and the banks may be less inclined to make investments in the system.

12 Free-entry equilibrium

Fixed-proportions division of B

Let fixed system costs be given by B and assume that this cost is covered by issuers only. Assume that issuers compete in quantities and each take a , the interchange fee, as given. Assume initially that B is divided equally between s issuers. Then, under the linear representation of demand introduced in Section 11, i.e., with inverse demand given by eqs. (33) and (34), the profit of a single issuer x is given by

$$\pi_{1x}(s) = ((L - c_1 + a)/(s + 1))^2(M - c_2 - a) - B/s \quad (44)$$

This is the standard expression for firm profit under Cournot competition with linear demand, constant marginal costs and entry cost B/s , except for the term that represents the network effect from the other side of the market, $M - c_2 - a$. The equilibrium will be characterized by

$$\begin{aligned} p_1 &= (L + s(c_1 - a))/(s + 1) \\ \pi_{1x}(s) &\geq 0 \\ \pi_{1x}(s + 1) &< 0 \end{aligned} \quad (45)$$

i.e., the standard free-entry equilibrium condition under Cournot competition. Since profits will be close to zero for any a , the issuers (or the acquirers) are not likely to have strong opinions about its level.

Per-transaction division of B

If the fixed cost B is instead allocated as a variable fee, the marginal cost of the issuer will be $c_1 - a + B/(n_1 n_2)$. As before, the acquirers marginal cost will be $c_2 + a$. In equilibrium, prices will equal marginal costs. Therefore, from eqs. (33) and (34) we have that

$$\begin{aligned} c_1 - a + B/(n_1 n_2) &= L - n_1 \\ c_2 + a &= M - n_2 \end{aligned}$$

or

$$\begin{aligned} n_1 &= L - c_1 + a - B/(n_1 n_2) \\ n_2 &= M - c_2 - a \end{aligned}$$

Solving for n_1 and substituting into eq. (33), we find that¹¹

$$p_1 = \frac{L + c_1 - a}{2} - \sqrt{\left(\frac{L - c_1 + a}{2}\right)^2 - \frac{B}{M - c_2 - a}} \quad (46)$$

The following proposition shows that prices will be at least as high if B is divided in fixed proportions, compared to a per-transactions division.

¹¹There are two solutions in the equation with $B/(n_1 n_2)$. Of these, the smaller value will be the equilibrium. At the larger value, although profits would be zero, an entrant would make a profit, as would a firm that expanded capacity.

Proposition 6. Under free entry and quantity competition, for linear demand, for a given interchange fee¹² and in the absence of strategic reasons for merchants to accept cards, prices will be at least as high in the issuing market when central system costs are divided between issuers in fixed proportions, compared to when system costs are allocated on a per-transaction basis.

Proof:

Under fixed-proportion division and free entry, each issuer's profit will be

$$\pi_i^{FP} = \frac{(p_1(s) - c_1 + a)\phi_1^{-1}(p_1(s))}{s}n_2 - \frac{B}{s} \quad (47)$$

where $\phi_1^{-1}(p_1)$ is the demand function representing the issuing side of the market. From free entry, it follows that

$$\pi_i^{FP} \geq 0 \quad (48)$$

Under a per-transaction division and free entry, total industry (issuer) profit will be

$$\Pi^{PT} = (p_1 - c_1 + a)\phi_1^{-1}(p_1)n_2 - B \quad (49)$$

From free entry, it follows that $\Pi^{PT} = 0$. Multiplying eq. (47) with s yields the industry profit under a fixed-proportions division

$$\Pi^{FP} = (p_1 - c_1 + a)\phi_1^{-1}(p_1)n_2 - B \geq 0$$

Since at the competitive price level, defined by eq. (49), a higher profit implies higher prices, it follows that prices under a per-transaction division will be no higher than prices under a fixed-proportion division - and sometimes lower. \diamond

According to the proposition, prices will be weakly lower when fixed costs are divided on a per-transaction basis. Notice also that the number of firms under a per-transaction division will be so high that each firm is a price taker; in theory, an infinite number.

13 Cournot competition without entry

When the number of firms is fixed, the above results are reversed: allocating costs in fixed proportions rather than in variable proportions will now result in *lower* prices, as demonstrated by the following proposition.

¹²Remember that under perfect competition, the banks have no reason to prefer a particular level of the interchange fee, since they will earn zero profit for any interchange fee.

Proposition 7. In a duopoly without entry, with firms that compete in quantities, with linear demand, for a given interchange fee and in the absence of strategic reasons for merchants to accept cards, prices will be higher on the issuing side of the market if fixed system costs are allocated in proportion to the number of transactions, rather than in fixed proportions.

Proof.

Demand is given by eqs. (33) and (34). If the fixed system costs are shared in equal and fixed proportions between the two firms, the equilibrium price on the issuing side of the market will be $p_1 = (L + 2(c_1 - a))/(3)$ (Cf. eq. (45), where $s = 2$ has been used). The profit of each of the two firms is given by

$$\pi_{11} = \pi_{12} = ((L - c_1 + a)/(3))^2(M - c_2 - a) - B/2 \quad (50)$$

The first-order condition for profit maximization corresponding to this expression will be

$$(L - 3n_{1x} - c_1 + a)(M - c_2 - a) = 0 \quad (51)$$

where n_{1i} is the quantity chosen by firm $i = 1, 2$ in the issuing side. By assumption, $M > c_2 + a$. It follows that $L - 3n_{1x} - c_1 + a = 0$ or $n_{1x}^{FP} = (L - c_1 + a)/3$. If, instead, the issuers pay for the system costs in proportion to their share of the total number of transactions, issuer 1's profit will be

$$\pi_{11} = (L - n_{11} - n_{12} - c_1 + a)n_{11}(M - c_2 - a) - \frac{n_{11}}{n_{11} + n_{12}}B \quad (52)$$

where $n_{1i} = 1, 2$ is still firm i 's quantity choice and where $n_{11} + n_{12} = n_1$. Differentiating with respect to n_{11} and using the symmetry of the firms, the first-order condition can be written

$$(L - 3n_{1x} - c_1 + a)(M - c_2 - a) - \frac{B}{4n_{1x}} = 0 \quad (53)$$

Comparing eq. (51) with eq. (53), we see that prices on the issuing side of the market will be higher under a per-transaction allocation of system costs. This is so, since $B/4n_{1x} > 0$ and, therefore, $L - 3n_{1x} - c_1 + a > 0$. This, in turn, implies that $n_{1x}^{PT} < (L - c_1 + a)/3$. Finally, since $n_{1x}^{PT} < n_{1x}^{FP}$, it follows that prices are lower under a fixed-proportions division of fixed system costs. \diamond

The intuition behind the above result is simple. When fixed costs are allocated on a per-transaction basis, the marginal cost of the two issuers will increase. Under Cournot competition and with linear demand, this will raise

equilibrium prices. If, instead, costs are divided in fixed proportions, central system costs will not influence the issuers' marginal costs and, consequently, final-customer prices will be lower in equilibrium.

Endogenous interchange fee

Possibly, the above result is an artifact of the assumption of an exogenously fixed interchange fee. In order to shed light in this issue, the interchange fee can be made endogenous. Assume that the two duopolists first fix the interchange fee and then compete in quantities. Since they are identical, the optimal interchange fee will be equal for the two firms.

Differentiating eq. (50) - the profit under duopoly and a fixed-proportion division of the system costs - with respect to a and solving for optimal a gives the same solution as the one that was derived for the case of no system costs, i.e., $a = \frac{2}{3}M - \frac{1}{3}L + \frac{1}{3}c_1 - \frac{2}{3}c_2$. Hence, when the fixed costs of the central system are allocated in fixed proportions, not only will pricing towards cardholders be unaffected by the costs for the central system, but so will the firms' joint choice of interchange fee and, consequently, price in the acquiring market. As for the issuers' willingness to invest in order to reduce marginal costs, i.e., to increase B in order to reduce c_1 , it is apparent from eq. (50) that they will have at least *some* incentives. A lower c_1 will increase profit although, of course, a higher B will work in the opposite direction. In general, since part of the benefit of a reduced c_1 will accrue to the cardholders, the issuers will tend to invest too little.

Unfortunately, the analysis becomes much more complex when the fixed system costs are allocated in variable proportions. Eq. (52) is not expressed in terms of exogenous parameters. Solving for the quantity choices of the issuers and substituting back into the equations will result in complicated expressions, that are not easily analysed. However, considering the results of previous sections, it appears reasonable to conjecture that the issuers will try to reduce their subjective marginal cost by raising the interchange fee. This, in turn, will raise prices in the acquiring market. Possibly, allocating central system costs in variable proportions will raise prices on both sides of the market, although prices on the issuing side will not rise as much as when the interchange fee is exogenous. If this conjecture is accurate, welfare is higher under a fixed-proportions allocation of system costs than under a per-transaction allocation, irrespective of whether the interchange fee is endogenous or exogenous.

14 Conclusions

Network effects in two-sided payment markets will result in quantities being too small (prices too high), also when the banks act as price takers. This is so, since the buyers do not (cannot) incorporate the positive external effect of buying another unit, while the sellers do not (cannot) incorporate the positive external effect of reducing the price. If the banks have market power, matters will be even worse - prices will then be even further from the social optimum.

In competitive markets, an interchange fee can potentially be used to improve

welfare. However, the banks will not have incentives to set the interchange fee at any particular level, since they will make zero profit regardless of its level.

In contrast, if the banks have market power, they will in general have an interest in setting a non-zero interchange fee. Issuing banks will want to set the interchange fee so as to extract profit from the acquiring side of the market, and vice versa. In addition, banks on both sides will have an incentive to set the interchange fee with an eye to system-wide network effects. If there is a monopoly on one side of the market, the welfare effect of introducing an interchange fee is indeterminate, given that it is set at the profit-maximizing level. The reason is not that there will be over-provision of card services (Cf. the results of Rochet and Tirole, 2002, referred to in Section 10), but that the market power on, say, the issuing side will now spill over to the acquiring side of the market.

From a policy perspective, there appears to be (at least) three possible regulatory strategies. First, the banks could be prohibited from using interchange fees. (Or, equivalently, the interchange fee can be restricted to be zero.) Second, the banks could be free to set interchange fees at the profit-maximizing level. Finally and third, the regulatory authorities (the bank regulator or the competition authorities) could try to approximate the optimal fee structure by allowing interchange fees at some intermediate level. The European Commission's competition directorate has opted for a version of the third strategy, when it decided to allow positive interchange fees - but only such fees as could be justified by costs. One way to interpret the decision is that the Commission accepted interchange fees that were no higher than the issuers' costs (corresponding to c_1 in this article).¹³ Indirectly, a similar regulatory regime has been established in the US, although this is the result of an out-of-court settlement following private litigation.¹⁴

From a welfare point-of-view, the optimal interchange fee is not related to the banks' cost structure (although the EU Commission in its decision has linked the two). This is under the assumption that the issuers and the acquirers carry the marginal costs that a transaction gives rise to on the respective sides. (In the analysis, I have abstracted from the fact that it may not always be apparent which side of the market that causes a given cost. By their very nature, a transaction involves both sides of the market.)

However, in most cases there will also be fixed central system costs. There is no single "correct" way to allocate these costs between issuers and acquires. In addition, the fixed costs can be distributed between the banks in fixed proportions, e.g., equal proportions between all active banks, but they can also be distributed in proportion to the number of transactions. In the latter case, but not in the former, they will influence the marginal costs of the banks, which in turn will tend to raise prices. On the other hand, if entry is possible, entry will be much easier if an entrant will not have to pay a large fixed cost in order to use the system.

¹³See the EU Commissions decision 24.07.2002 in the *Visa International* case, COMP/29.373, published in the Official Journal L 318, 22.11.2002, p. 17-36 .

¹⁴See the Economist, May 3, 2003, p. 66-67.

Under free entry and for linear demand, the entry effect dominates: prices will be at least as low (and sometimes lower), if costs are distributed in proportion to the number of transactions. On the other hand, if there is *no* entry, costs will be *higher* under proportional cost distribution.

Another concern is the banks' willingness to incur fixed costs in the central system in order to reduce marginal processing costs; this willingness will be greater under a proportional cost-distribution scheme.

The policy conclusion to be drawn from the analysis of this article is that if entry barriers are not too high, fixed costs should be allocated in proportion to the number of transactions. This will facilitate entry which, in turn, will put a downward pressure on prices. In addition, the banks will have better incentives to make investments in technology that reduces marginal costs.

On the other hand, if entry barriers are high, or if entry is impossible, the welfare consequences of allocating fixed costs in proportion to the number of transactions appears to be more problematic. This method of cost allocation will now tend to raise prices, although there will still be a positive incentive effect on investments.

A limitation of the present study is that prices are assumed to be linear. In practice, the pricing schemes that cardholders and merchants face are more complex, involving, e.g., fixed annual fees. Consequently, the effect of non-linear prices in two-sided markets with fixed system costs would be an interesting field for further studies.

15 References

Armstrong, Mark 2004, *Competition in Two-Sided Markets*, mimeo, University College London.

Baxter, W.P., 1983, Bank Interchange of Transactional Paper: Legal Perspectives, *Journal of Law and Economics*, 26, 541-588.

Chakravorti, Sujit, 2003, Theory of Credit Card Networks: A Survey of the Literature, *Review of Network Economics*, 2, 50-68.

Chakravorti, Sujit and Roberto Roson, 2004, *Platform Competition in Two-Sided Markets: The Case of Payment Networks*, mimeo, Federal Reserve Bank of Chicago and Department of Economics, Università Ca' Foscari di Venezia.

Gans, J.S. and S.P. King, 2003, The Neutrality of Interchange Fees in the Payment Systems, *Topics in Economic Analysis & Politics*, 3, 1, Article 1. (See <http://www.bepress.com>.)

Guthrie, Graeme and Julian Wright, 2003, *Competing Payment Schemes*, Working Paper No. 245, University of Auckland.

Katz, Michael, 2001, *Reform of Credit Card Schemes in Australia II*, Reserve Bank of Australia, Sydney, Australia.

Manenti and Somma, 2003, *Plastic Clashes: Competition among Closed and Open Systems in the Credit Card Industry*, mimeo, Università di Padova and Università degli Studi di Bari.

Rochet, Jean-Charles and Jean Tirole, 2002, Cooperation Among Competitors: Some Economics of Payment Card Associations, *Rand Journal of Economics*, 33, 549-570.

Rochet and Tirole, 2003, Platform Competition in Two-Sided Markets, *Journal of the European Economic Association*, 1, 990-1029.

Schmalensee, Richard, 2002, Payment Systems and Interchange Fees, *Journal of Industrial Economics*, 50, 103-122.

Wright, Julian, 2004, The Determinants of Optimal Interchange Fees in Payment Systems, *Journal of Industrial Economics*, 52, 1-26.